International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Special Issue: | Apr 2023 www.irjet.net p-ISSN: 2395-0072

International Conference on Trends in Engineering & Technology- 2023 (ICRTET)
Organised by: VSM College of Engineering, Ramachandrapuram

# DATA MINING MODEL BASED ON DISCRIMINATIONOF AMBIGUOUS BEHAVIOR IN WATER CONSUMPTION

**[1] M. Naga Jyothi Assistant Professor, [2] K. Surekha Associate Professor, [3]P.Ramya Assistant Professor**

*Dept of Computer Science, V.S.M. College of Engineering Ramachandrapuram*

---

**ABSTRACT**

Water supply companies and government struggle greatly with drinking water fraud. The majority of non-technical losses are caused by this activity, which has a considerable negative financial impact. Recentresearch has placed a lot of emphasis on developingreliable criteria for identifying dishonest behavior's. Waterdelivery firms have access to sophisticated data mining techniques that can be used to identify fraud andcut losses. The purpose of this study istoinvestigate the use of three classification algorithms (SVM,KNN,ANN) to identify potentialwater scam victims. This study's primary goal is to help the Yarmouk Water Company (YWC) in Irbid, Jordan, fill its revenue gap. Known anomalous activity is exposed by the SVM-based approach using customerload profile characteristics.

## I. INTRODUCTION

Water theft reduces the organization's revenue and makes it more difficult to give water to all citizens equally. Despite the fact that the MOI is a nonprofit organisation tasked withproviding water to all residents, maintaining a balance between costs and income is essential to providing an equitable water supply to everyone. A false consumption amount is produced by fraudulent water use since it is crucial to properly quantify the energyused in order to increase profits. Around 8000 buildings in the city, according to the MOI, receive water without the aid of a water metre, which is what the MOI uses to determine each client's monthly usage.

There are advantages and disadvantages to each strategy that has been developed by the many peoplewho are affected by this fraudulent behaviour and who want to stop it in addition to recovering from the losses brought on by it. We developed a data mining model to get around some drawbacks and increase the advantages. In comparison to SVM, other approaches like KNN and ANN are less accurate and perform worse. The SVM performs and is accurate really well. The SVM achieves an intelligent detection rate of 80% and a detection rate for random data of 1–10%. Depending on the customer's activities and behaviour, the customer profile can be altered. According on the client's choices, this strategy divides the customer data into three categories: monthly, seasonally, and yearly. To determine the client's behaviour, past customer data is gathered and put to the test. This study contributes to the provision of customer load profiles according to water consumption patterns. The primary goal of this study is to distinguish between genuine customers and scam customers. Customers' water consumption is tracked in this system.

## II.PROCEDURE

The primary goals of this study are to

1. Help the MOI reduce its NTL's for the water distribution sector.

2. Consider the challenges of dealing with only 20% of the available customer load profile data, as well as the ability to detect and identify NTL actions using data mining classificationalgorithms.

3. Criminal activities In order for the intelligent model developed in this research study to forecast suspicious customers and help them uncover fraud activities, MOI DWTC staff manually and randomly inspect clients on-site. Recent data from the water distribution agency show that the financial losses brought on by water use are due to a significant discrepancy between the city's water well production and water use, as seen in Fig 2. The difference increased to 15 in2011.

This method is used to categorize clients based on their conduct. The technique determines whether a customer is fraudulent or not by comparing their data with information that has already been saved. When classifying data, SVM,KNN and ANN classification models are used because of their high accuracy and superior performance.

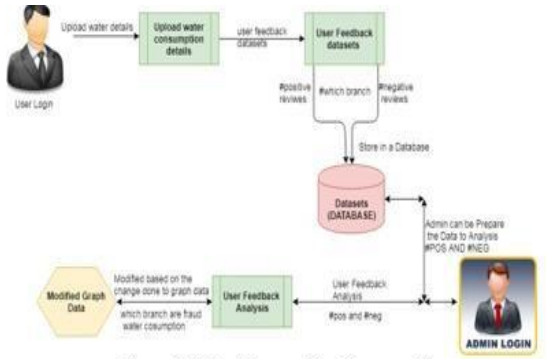



Figure 1 Utility of linear and nonlinear model

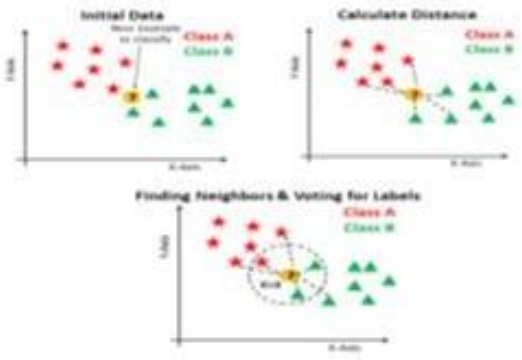
Customers need to register with the system in order to purchase water from agencies. You can upload data about your water consumption after registering. Field executives use this technique to gather client feedback andmonitor branch-specific restrictions. Administrators can arrange the data to examine both positive and negative feedback. Based on the user's comments, fraud details are blocked,prohibiting them from receiving any more water. The problem's root cause can be found by the admin, who can find the fraudulent consumers. This will give a clear picture of the dataset'scurrent and historical state.

## A. Algorithm

In this study, experiments are conducted using the K-Nearest Neighbour (KNN), Artificial Neural Network (ANN) and SVM classification algorithms, which are applied to a data set of water customers4 obtained from theExecutives. The datasets are evaluated to determine their accuracy. Each method has been applied to the trainingdataset, and its accuracy performance has been assessed together with the forecasting performed on the testing dataset. One of the most popular and widely applied open source data mining methods worldwide. It offers a pleasant user interface, whereconfigurations for analyses are made inthe process view. It employs a modularconcept where the analysis process applies the appropriate operators. These operators can interface with other operators via I/O ports to receive input data or to transfer data and models created by one operator to the next. In this approach, data flow is produced throughout the analytical process. Forecasts are produced using the K nearest neighbouring results (K-Nearest Neighbour).5 We must therefore develop a measure formeasuring the distance between the query point and the instances in theexamples sample in order to make predictions with KNN and ANN.

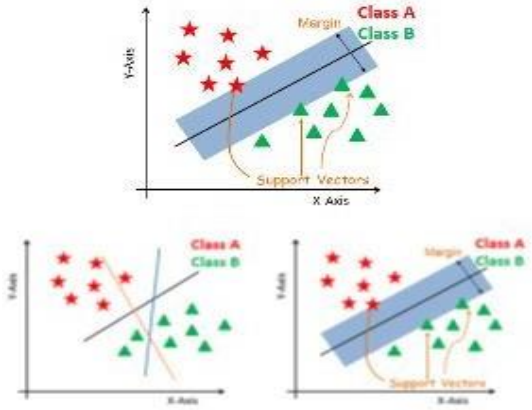
1. KNN Algorithm

(Image)[11]




2. SVM Algorithm

(Image)[10]




3. ANN Algorithm